Server Load Balancer

Best Practices

MORE THAN JUST CLOUD |

Best Practices

How to use guaranteed-performance instances?

What are guaranteed-performance instances?

Performance metrics, such as Max Connection, CPS, and QPS, are included in the guaranteedperformance instance SLA. In contrast, shared-performance instances do not provide the performance guarantees. The Server Load Balancer resources are shared among the sharedperformance instances.

All instances are shared-performance instances before Alibaba launches guaranteed-performance instances. You can view the instance type on the console.

You can hover your mouse pointer to the green icon of the target guaranteed-performance instance to view the performance metrics, as shown in the following figure.

 ID/Name Ib- (Name) 	Zone cn-hangzhou- f(Master)	IP Address(All) ~ 47.98.17.89(Public IP)	Status	Network(All) - Classic Network	Port/Health Check	Backend Server	Specification Guaranteed- Performance	Method(All) +	Method(All) + Pay-As-You-Go	Actions Manage More
(None) 🖌	e(Slave)				comgareacomgare	conigareaconigare	slb.s1.small	CPS: 3000 QPS: 1000	ated	

The following are three key metrics of guaranteed-performance instances:

Max Connection

The maximum number of connections to a SLB instance. When the maximum number of connections reaches the limits of the specification, the new connection will be dropped.

Connection Per Second (CPS)

The rate at which a new connection is established per second. When the CPS reaches the limits of the specification, the new connection will be dropped.

Query Per Second (QPS)

The number of HTTP/HTTPS queries/requests that can be processed per second, which is specific to layer-7 listeners. When the QPS reaches the limits of the specification, the new connection will be dropped.

Alibaba Cloud Server Load Balancer provides the following specifications for guaranteedperformance instances:

Specification		Max Connection	CPS	QPS	
Specification 1	Small I (slb.s1.small)	5000	3000	1000	
Specification 2	Standard I (slb.s2.small)	50000	5000	5000	
Specification 3	Standard II (slb.s2.medium)	100000	10000	10000	
Specification 4	Higher I (slb.s3.small)	200000	20000	20000	
Specification 5	Higher II (slb.s3.medium)	500000	50000	30000	
Specification 6	Super I (slb.s3.large)	1000000	100000	50000	

How to choose specifications for guaranteedperformance instances?

Choose the specification according to your service types, and the overall principle is as follows:

The key factor for layer-4 listeners is the number of concurrent connections of TCP keepalive connections, then the max connection is considered as the key metric. Depending on the business scenarios, estimate the maximum number of concurrent connections and select the appropriate specification.

The key factor for layer-4 listeners is QPS performance. QPS determines the throughput of a layer-7 application system. Similarly, you also need to estimate the QPS based on experience. After the initial selection of a specification, you can adjust the specification during business stress test and real test.

Use other monitoring metrics introduced by guaranteed-performance instances to check the traffic trend, peak traffic, and so on for more accurate selection. For more information, see Monitoring data.





Limits on the configuration change of a guaranteedperformance instance

You can change the configuration of a guaranteed-performance instance, as shown in the following figure.

Server Load Balancer ID/Name	Zone	IP Address(All) +	Status	Network(All) +	Port/Health Check	Backend Server	Instance Specification	Bandwidth Billing Method(All) +	Billing Method(All) +	Actions
None)	cn-hangzhou- f(Master) cn-hangzhou- e(Slave)	47 89(Public IP)	Running	Classic Network	Not ConfiguredConfigure	Not ConfiguredConfigure	Guaranteed- Performance Instance slb.s1.small ()	Pay by Traffic	Pay-As-Y 2018-01- 16:02:5* Created	ou-Go -29 Mai Start	nage More+
b- ab8b4478a00ec11e89	cn-hangzhou- b(Master) C-hangzhou- d(Slave)	120145(Public IP)	Running	Classic Network	TCP: 80	Not ConfiguredConfigure	Shared- Performance Instance	Pay by Traffic	Pay-As- 2018-0 17:55:1 Created	Stop Release Edit Tags Change Co	nfiguration

instance name: lb-1/ mh										
Billing item : Configuration fee+Traffic fee		+Traffic fee	Instance Spec : Small I (slb.s1.sm	II) Primary zone : c	n-hangzhou-f	Backup zone : cn-hangzho	u-e			
andwidth	h : By traffic		Instance type : Internet	Region : China I	ast 1	slb rentalfee : Yes				
nti-DDos	s : Enabled		Zone type : Multi-zone							
Configu	ration upgrade									
Configu	ration upgrade									
Configu	ration upgrade									
Configu	ration upgrade									
Configu	Instance type	Internet								
e fype	Instance type	Internet	•							
tance type	Instance type	Internet Small I (slb s1	imali) 👻							
configu	Instance type	Internet Small I (slb.s1.	mali) 👻							
Configure type	Instance type	Internet Small I (slb.s1. Max connection:	mali) •							
Configu	Instance type	Internet Small I (slb.s1. Max connection:	mali)							

Note: Some instances may be in old clusters because of historical inventory. These instances need to be migrated when they are changed to guaranteed-performance instances, thereby the corresponding Server Load Balancer service will be interrupted for 10s-30s. We recommend that you do this change when the traffic is low, or use **GSLB** to do global load balancing first and then change the configuration.

Warning
Please note that the Server Load Balancer service will be interrupted for 10s- 30s, if you change a performance-shared instance to the performance- guaranteed instance. We recommend that you do this change when the traffic is low, or use DNS to switch the service to another Sever Load Balancer instance before changing the instance specification. The changes to the bandwidth and billing method will not impact your business.
I am aware of the risk, continue Cancel

Pricing of guaranteed-performance instances

Different specification fees are charged on guaranteed-performance instances with different specifications. Alibaba Cloud also provides a free specification that can meet the demands of most shared-performance instance users.

Note: In addition to the specification fee, you also need to pay the configuration fee and traffic fee for an instance. For more information, see **Billing**.

How to handle the original shared-performance instances?

The original shared-performance instances will not be automatically upgraded to guaranteedperformance instances and also will not be charged for the specification fee.

You can manually upgrade them to guaranteed-performance instances. After upgrading, you will be charged for the specification fee accordingly.

Note: Some of the shared-performance instances may be deployed in an old cluster. When upgrading these instances to guaranteed-performance instances, a service interruption of 10-30 seconds may occur during the migration of the instances. We recommend that you upgrade these instances in a low traffic period. The upgrading of the guaranteed-performance instances has no impact on the services.

Why sometimes guaranteed-performance instances cannot reach the performance limit defined in the specification?

It is because of the short board principle.

Guaranteed-performance instances do not guarantee that the three metrics can reach the specification limits at the same time. That is, when one metric first reaches its limit, limitation is triggered.

For example, you have purchased a guaranteed-performance instance of specification higher I (slb.s3.small). When the QPS of the instance reaches 20,000 but the number of maximum connections does not reach 200,000, the new connections are still dropped because the QPS has reached the limitation.

Note: If you use Anti-DDoS Pro so that there are few source IPs, the HTTPS traffic may be unbalanced. You can open a ticket to solve the problem.

Why sometimes the performance of a guaranteedperformance instance is worse than that of a sharedperformance instance?

For a shared-performance instance, all the resources are shared. Its performance cannot be guaranteed when the traffic load is high. However, a guaranteed-performance instance can guarantee the performance at any time.

When can I use API to create and modify guaranteedperformance instances?

Now the creation and modification of guaranteed-performance instances is not supported by Server

Load Balancer API. Check your registered email account and Alibaba Cloud website for further notifications.

Can I still buy shared-performance instances?

Yes. However, shared-performance instances will be unavailable in the future. Check the registered email account and Alibaba Cloud website for further notifications.

Configure cookie in the backend server

Server Load Balancer provides session persistence function. With session persistence enabled, Server Load Balancer can distribute requests from the same client to the same backend server during the session period.

For layer-4 listeners, session persistence is based on the IP address. The listener of Server Load Balancer forwards requests from the same IP address to the same backend server.

For layer-7 listeners, session persistence is based on cookies. If you choose the **Rewrite Cookie** method, you can set the **Cookie Name** as name, and set the key of vip.a.com 's cookie as name on the backend server.

Hide Advanced Options	
Obtain Real IP:	Enable(Default)
Session Persistence:	Enable HTTP session persistence is based on cookie.
Cookie Handling:	Rewrite Cookie
Cookie Name:*	name
	periods (.) or commas (,).

Follow the instructions in this section to set cookies on a backend server.

Apache

Open the httpd.conf file and make sure that the following line is not commented.

LoadModule usertrack_module modules/mod_usertrack.so

Add the following configurations in the VirtualHost file.

CookieName name CookieExpires "1 days" CookieStyle Cookie CookieTracking on

Nginx

Configure the configuration file as follows.

```
server {
  listen 8080;
  server_name wqwq.example.com;
  location / {
   add_header Set-Cookie name=xxxx;
  root html;
  index index.html index.htm;
  }
}
```

Lighttpd

Configure the configuration file as follows.

```
server.modules = ( "mod_setenv" )
$HTTP["host"] == "test.example.com" {
server.document-root = "/var/www/html/"
setenv.add-response-header = ( "Set-Cookie" => "name=XXXXXX" )
}
```

Use Open API to configure Server Load Balancer

In this tutorial, the request parameters are included in the request URL, and the URL does not include common parameters. For more information, see API overview.

Note: To increase readability, the parameter values of the request URL in this example are not URL-encoded.

Prerequisites

You have created 2 ECS instances and granted access to their SSH and Web ports.

Procedure

Call CreateLoadBalancer interface to create a Server Load Balancer instance.

Request:

https://slb.aliyuncs.com/?Action=CreateLoadBalancer&RegionId=cn-hangzhou-dg-a01

Response:

```
{
    "RequestId":"3DE96B24-E2AB-4DFA-9910-1AADD60E13A5",
    "LoadBalancerId":"LoadBalancerId",
    "Address":"SLBIPAddress"
}
```

Call CreateLoadBalancerHttpListener interface to create a HTTP listener, of which the port is 80, for the Server Load Balancer instance.

Request:

https://slb.aliyuncs.com/?Action=CreateLoadBalancerHttpListener&LoadBalancerId=LoadBa lancerId&ListenerPort=80&BackendServerPort=80&ListenerStatus=active

Call SetLoadBalancerStatus interface to active the Server Load Balancer instance.

Request:

```
https://slb.aliyuncs.com/?Action=SetLoadBalancerStatus&LoadBalancerId=LoadBalancerId&LoadBalancerStatus=active
```

Call AddBackendServers interface to add an ECS instance to backend servers.

Request:

```
https://slb.aliyuncs.com/?Action=AddBackendServers&LoadBalancerId=LoadBalancerId&Ba
ckendServers=[{"ServerId":"ECS1InstanceID"}]
```

Response:

```
{
"RequestId" : "FA2F2172-63F2-409D-927C-86BD1D536F13",
"LoadBalancerId" : "LoadBalancerId",
"BackendServers" : {
"BackendServer" : [
{
"ServerId" : "ECS1InstanceId",
"Weight" : 100
}
]
```

Call AddBackendServers interface again to add an ECS instance to backend servers.

Request:

```
https://slb.aliyuncs.com/?Action=AddBackendServers&LoadBalancerId=LoadBalancerId&Ba
ckendServers=[{"ServerId":"ECS2InstanceID"}]
```

Response:

```
{
"RequestId" : "C61FAD0A-2E87-4D0C-80B0-95AB758FCA70",
"LoadBalancerId" : "LoadBalancerId",
"BackendServers" : {
"BackendServer" : [
{
"ServerId" : "ECS1InstanceId",
"Weight" : 100
},
{
"ServerId" : "ECS2InstanceId",
"Weight" : 100
}
]
}
}
```

Call DescribeLoadBalancerAttribute interface to view the configuration of the Server Load Balancer instance.

Request:

https://slb.aliyuncs.com/?Action=DescribeLoadBalancerAttribute&LoadBalancerId=LoadBal ancerId

Response:

{

```
"RequestId" : "4747E9AE-ADFD-412D-B523-C1CBD45A2154",
"LoadBalancerId" : "LoadBalancerId",
"Address" : "SLBIPAddress",
"IsPublicAddress" : "true",
"ListenerPorts" : {
"ListenerPort" : [
80
1
},
"BackendServers" : {
"BackendServer" : [
{
"ServerId" : "ECS1InstanceId",
"Weight" : 100
},
{
"ServerId" : "ECS2InstanceId",
"Weight" : 100
}
]
}
}
```

Use your browser to access the IP address of the Server Load Balancer instance to verify whether the service is working.

Remove backend ECS

Directly removing backend ECS instances from a Server Load Balancer instance may cause service interruption. We recommend setting the weight of an ECS instance to zero first, and then remove it when no traffic is distributed to it.

Log on Server Load Balancer console.

Choose a region and then click the ID of the target Server Load Balancer instance.

In the left-side navigation pane, click Server > Backend Server.

If the ECS instance is added to a server group, click **VServer Group** or **Master-Slave Server Group** accordingly.

Hover the mouse pointer to the weight of the target ECS instance and then set the value to **0**.

<	CCS- Return to Server Load Balancer List.	rictions and Notes										
Details Listener	Load Balancer Server Pool Region : China East 1 (Hangzhou) Zone : cn-hangzhou-f (Master) /cn-hangzhou-e (Slave) 🛛											
▼ Server	Servers Added Servers Not Added											
Backend Server	Instance Name v Enter the instance name of the ECS server. Search	C Refresh										
VServer Group	ECS Instance ID/Name Zone Public/Internal IP Address Status(All) - Network Type(All) - Health Check Weight	Action										
Master-Slave Server	Hob12gcip/immcrtiaavd drey11/9222/b486 cn- hangzhou-f 118 121 (Elsatic) © Running VPC (vpc-bp15p1s/mx3fg2r5suredry) Normal 100	Remove										

When no traffic is distributed to the ECS instance, click **Remove** to remove it from the backend server pool.

Troubleshoot

If there are ongoing service requests sent to the ECS instance after removing it from the backend server pool, check the following:

Whether the ECS instance is added to backend server pools of other Server Load Balancer instances.

You can use the ECS instance ID to filter Server Load Balancer instances that the ECS instance is added to.

Instance Management												
China North 1 (Qingdao)	China North 2	2 (Beijing)	China North 3 (Zha	ingjiakou)	China East 1 (Hangzhou)	China East 2 (Shang	hai) China :	South 1 (Shenz	nen)		
Hong Kong Asia Pacifi	Singapore	Asia Pacific SE 2 ((Sydney)	US East 1 (Virg	inia) US V	Vest 1 (Silicon Valley)	Middle East	1 (Dubai)				
Germany 1 (Frankfurt)	Germany 1 (Frankfurt)											
Create Server								Load Bala	ncer			
Backend Server ID	٣	i-bp12gicjp	6mr		Search	h 🔊	ag				<u>)</u>	٥
Server Load Balancer ID/Name	Zone	IP Address((All) Status	Network(All	l) Port/Health	Check	Backend Server	Instance Spec	Bandwidth Billing Method(All)	Billing Method(All)		Action
b-1udstfm38 acs-slb-cfe9117922	cn-hangzhou- f(Master) cn-hangzhou-	120 Public IP	Running	Classic Network	TCP: 80	Normal	cfe91179	Shared- Performance Instance	Pay by Traffic	Pay-As-You- Go 2017-09-16 14:27:43	Manage	More≠

Log on to the ECS instance, run the netstat command to check whether the ECS instance is deployed with public services.

Windows: Run netstat -ano to view all open ports on the instance.

Linux: Run this command to view all open ports on the instance or use other parameters of the netstat command.

			:~# ne	tstat -ar	10			
Activ	e Interr	net conr	nections (serve	rs and es	stablished)			
Proto	Recv-Q	Send-Q	Local Address		Foreign Ad	dress	State	Timer
tcp	Θ	Θ	0.0.0.0:22		0.0.0.0:*		LISTEN	off (0.00/0/0)
tcp	Θ	Θ	0.0.0.0:111		0.0.0.0:*		LISTEN	off (0.00/0/0)
tcp	Θ	Θ	172.16.	:42285		`:80	ESTABLISHED	off (0.00/0/0)
tcp	Θ	428	172.16.	22		:44	332 ESTABLISHED	on (0.16/0/0)
tcn6	0	Θ	:::111		*		LISTEN	off (0.00/0/0)
udo	õ	õ	0.0.0.0:42947		0.0.0.0:*		Eroten	off $(0, 00/0/0)$
udp	õ	õ	0.0.0.0.68		0.0.0.0.*			off (0.00/0/0)
udp	6	0	0 0 0 0 0 111		0.0.0.*			off (0.00/0/0)
udp	6	0	0.0.0.0.0.111		0.0.0.0.*			off $(0.00/0/0)$
uup	0	0	0.0.0.0:027	100	0.0.0.0.*			
uap	U	0	1/2.10.	:123	0.0.0.0:*			011 (0.00/0/0)
udp	Θ	Θ	127.0.0.1:123		0.0.0.0:*			off (0.00/0/0)
udp	Θ	Θ	0.0.0.0:123		0.0.0.0:*			off (0.00/0/0)
udp6	Θ	Θ	:::111		:::*			off (0.00/0/0)
udp6	Θ	Θ	:::627		:::*			off (0.00/0/0)
udp6	Θ	Θ	:::123		:::*			off (0.00/0/0)
udp6	Θ	Θ	:::1275		:::*			off (0.00/0/0)
Activ	e UNIX d	domain's	ockets (server	s and est	tablished)			
Proto	RefCnt	Flags	Type	State	I-No	de Pat	th	
unix	2	[]	DGRAM		7689	/r	un/svstemd/shutdow	nd
unix	7	[]	DGRAM		7691	/r	un/systemd/journal,	/dev-log

Obtain the real IP address of the client

Introduction to the function of obtaining IP address

Alibaba Cloud Server Load Balancer provides the function of obtaining the real IP address of the client and this function is enabled by default.

For the Layer-4 load balancing service (TCP protocol), listeners distribute client requests to backend ECS servers without modifying the request headers. Therefore, you can obtain the real IP address from the backend ECS servers without additional configurations.

For the Layer-7 load balancing service (HTTP/HTTPS protocol), you have to configure the application servers, and then use the X-Forwarded-For header to obtain the real IP addresses of the clients.

Note: For the HTTPS load balancing service, the SSL certificates are configured in frontend listeners, the backend still uses the HTTP protocol. Therefore, the configurations on application servers are the same for HTTP and HTTPS protocols.

Add Listener	
1.Listener Configurati	on > 2.Health Check Configuration > 3.Success
Frontend Protocol [Port] *	HTTP v : You can enter any port number from 1-65535.
Backend Protocol [Port] *	HTTP : You can enter any port number from 1-65535.
Peak Bandwidth:	Unlimited Configure You can set a peak bandwidth from 1-5000M. By default, the instances charged by traffic do not have peak bandwidth limit.
Scheduling Algorithm:	Weighted Roun 🔻
Use VServer Group:	0
Automatically Activate Listener after Creation:	Activated
Collapse – Advanced Options	
Obtain Real IP:	Activated(Default)
Session Persistence:	Close HTTP HTTP sticky sessions are based on cookies.

Configure web applications

This section introduces some common methods used to configure web applications.

Configure IIS7/IIS8

Download and extract the F5XForwardedFor.

Copy the F5XFFHttpModule.dll and F5XFFHttpModule.ini files from the extracted folder to a folder, such as C:\F5XForwardedFor\. Make sure that the IIS process has the write permission to this folder.

Open the IIS Manager, and then double-click the Modules function.

File View Help								
Connections	🧤 iZe	eu6akphal	o4juZ Hoi	me				
Start Page ▲ Start Page ▲ Call Application Page	Filter:		• 🔻 Go 🕞 🦕	Show All	Group by: Are	ea	-	^
⊳	<u></u>	Ð	0		404	2	*	
	Authentic	Compression	Default Document	Directory Browsing	Error Pages	Handler Mappings	HTTP Respon	
		j			0	D	2	
	Logging	MIME Types	Modules	Output Caching	Request Filtering	Server Certificates	Worker Processes	
	Manageme	ent						- ^
			88	88		*		
	Configurat Editor	Feature Delegation	IIS Manager Permissions	IIS Manager Users	Management Service	Shared Configurat		

Click Configure Native Modules, and then click Register.

	Configure Native Madula	. ? X		Actions
	Configure Native Module	s L.		Add Managed Module
Use this Web ser	Select one or more registered modules to enable:		the	View Ordered List
Group	UriCacheModule	Register		Help
Name	TokenCacheModule	Edit	ntry Type	
Anony			ocal	
Custon		Remove	ocal	
Default			ocal	
Directo			ocal	
HttpCa			ocal	
HttpLo			ocal	
Protoco			ocal	
Reques			ocal	
StaticC			ocal	
StaticFi			ocal	
	Or	Cancel		
	UK	Cancel		

Add the copied the .dll file.

Register Native Module ? ×	
Name: F5XForwardedFor_64	
Path: C:\Users\Administrator\Desktop\F5XForwardedFor\F5XForwardedFor 	
OK Cancel	

Add the ISAPI and CGI restrictions for the .dll file and set the restriction to Allowed.

Make sure that you have installed the ISAPI and CGI applications.

File View Help								
Connections	Use this feature to specify the ISAPI and CGI extensions that can run on the Web server.							
IZeubakphab4juZ (IZeubakphi Application Pools	N Group by No Grouping							
Application roots	oroup by. No orouping							
⊳õi Sites	Description	Restriction	Path					
	Active Server P	Allowed	%windir%\system32\inetsrv\asp.dll					
	х64	Allowed	C:\Users\Administrator\Desktop\F5XForwardedFor\F5					
	x86	Allowed	C:\Users\Administrator\Desktop\F5XForwardedFor\F5					

Restart the IIS Manager.

Configure Apache

Run the following command to install the mod_rpaf module.

wget http://stderr.net/apache/rpaf/download/mod_rpaf-0.6.tar.gz tar zxvf mod_rpaf-0.6.tar.gz cd mod_rpaf-0.6 /alidata/server/httpd/bin/apxs -i -c -n mod_rpaf-2.0.so mod_rpaf-2.0.c

Open the /alidata/server/httpd/conf/httpd.conf file and add the following information at the end of the content.

LoadModule rpaf_module modules/mod_rpaf-2.0.so RPAFenable On RPAFsethostname On RPAFproxy_ips IP_address RPAFheader X-Forwarded-For

RPAFproxy_ips: the IP address is not the IP address of the Server Load Balancer instance. Check the Apache log to find the IP address, usually both the two IP addresses are entered.

Run the following command to restart the Apache server.

/alidata/server/httpd/bin/apachectl restart

Configure Nginx

Run the following command to install http realip module.

wget http://nginx.org/download/nginx-1.0.12.tar.gz
tar zxvf nginx-1.0.12.tar.gz
cd nginx-1.0.12
./configure --user=www --group=www --prefix=/alidata/server/nginx --with-http_stub_status_module -without-http-cache --with-http_ssl_module --with-http_realip_module
make
make install
kill -USR2 `cat /alidata/server/nginx/logs/nginx.pid`
kill -QUIT `cat /alidata/server/nginx/logs/ nginx.pid.oldbin`

Run the following command to open the nginx.conf file.

vi /alidata/server/nginx/conf/nginx.conf

Find the following content and add the required information after it.

fastcgi connect_timeout 300; fastcgi send_timeout 300; fastcgi read_timeout 300; fastcgi buffer_size 64k; fastcgi buffers 4 64k; fastcgi busy_buffers_size 128k; fastcgi temp_file_write_size 128k;

The information to be added:

set_real_ip_from IP_address
real_ip_header X-Forwarded-For;

set_real_ip_from IP: the IP address is not the IP address of the Server Load Balancer instance. Check the Nginx log to find the IP address, usually both the two IP addresses are entered.

Run the following command to restart the Nginx server.

/alidata/server/nginx/sbin/nginx -s reload

How to forward same-domain requests to different servers

In this case, we use four ECSs deployed with Nginx servers as the example to demonstrate how to configure forwarding rules specified by domain name and URL, so as to fulfill traffic forwarding as shown in the following table.

Frontend request	Forward traffic to
www.aaa.com/tom	Server SLB_tom1 and server SBL_tom2
www.aaa.com/jerry	Server SLB_jerry1 and server SBL_jerry2

Instance ID/Name	Zone	IP Address	Status(All)	Network Type(All) +
i-bp1 and an	\$ China East 1 Zone F	4 21(Elastic IP 1 1 1 1 1 21(Elastic IP 1 1 1 1 1 21(Private IP Address)	• Running	VPC
i-bp://www.apy SLB_jerry2	\$ China East 1 Zone F	41.96.172.148(Elastic IP indiana) 172.16.33.32(Private IP Address)	• Running	VPC
i-be The second se	\$ China East 1 Zone F	1 K S2 119 04 (Elastic IP Accessed) 17 16 19 00(Private IP Address)	• Running	VPC
i-bpl according to the hot SLB_tom2 🖌	\$ China East 1 Zone F	47.06.169.125(Elastic IP Address) 172.04.30(Private IP Address)	• Running	VPC

Procedure

Create an Internet-facing SLB instance.

For details, see Create a server load balancer.

Resolve the domain name into the public IP of the SLB instance by using DNS.

For convenience, the public IP of the SLB instance is bound to domain name www.aaa.com in the host file in this case.

Create two VServer groups.

Locate the newly created instance in the Server Load Balancer console and click the instance ID to go to the **Instance Details** page.

In the left-side navigation pane, click **Server > VServer Group**.

Click Create VServer Group.

In the dialog box that appears, select the backend servers to be added and set ports and weights for them respectively. The ports for ECSs in the VServer group can be different.

In this case, enter **TOM** as the server group name, add server SLB_tom1 and server SBL_tom2 into the group, set the port number to 80, and keep the default weight value (100).



Repeat the preceding steps to add another VServer group named JERRY, which includes server SLB_jerry1 and server SBL_jerry2.

Add a listener.

In the left-side navigation pane, click Listeners, and click Add Listener.

Configure the listener. In this case, the listener is configured as follows:

- i. Frontend protocol [Port]: HTTP: 80
- ii. Backend protocol [Port]: HTTP: 80
- iii. Scheduling algorithm: Round-robin.
- iv. Keep the default values for other configuration items.

On the Listeners page, click More > Add Forwarding Rules.

	training_SLB	*Return to Server Load Balancer Lis	t							Restriction	s and Notes
1 u	steners									Add Listener	Refresh
	Front-end Protocol/Port	Backend Protocol/Port	Status	Forwarding Rules	Session Persistence	Health Check	Peak Bandwidth	Server Group			Actions
	HTTP: 80 Start Stop	HTTP:80 Delete	Running	Round Robin	Disable	Enable	No Limits	-	Configure Details	Add Forwarding Rule Activate Stop Delete Set Access Co Add Forwardin	ntrol

On the Forwarding rules page, click Add Forwarding Rules.

Configure three forwarding rules.

Add Forwarding Rules								
Rule Name	Domain Name	URL	VServer Group	Actions				
rule1	www.aaa.com	/jerry	JERRY	Delete				
rule2	www.aaa.com	/tom	ТОМ	Delete				
rule3	www.aaa.com		ТОМ	/ Delete				
	Add	Forwarding Rule]					
 * Domain name rule: The domain name can contain letters a-z, numbers 0-9, hyphens (-), and periods (.), and wildcard characters. The following two domain name formats are supported: Standard domain name: www.test.com Wildcard domain name: *.test.com. wildcard (*) must be the first character in the format of (*.) * URL rule: 								
URLs must be 2-80 characters in length. Only letters a-z, numbers 0-9, and characters '-' '/' '?' '%' '#' and '&' are allowed. URLs must be started with the character '/', but cannot be '/' alone.								
* At least one domain name rule or URL rule is required.								
			Confirm	Cancel				

Test:

Enter www.aaa.com/jerry in the browser and the following result is returned.



Enter www.aaa.com/tom in the browser and the following result is returned.



Enter www.aaa.com in the browser and the following result is returned.

